J. Mol. Biol. (1992) 227, 381–388

By-passing Immunisation

Human Antibodies from Synthetic Repertoires of Germline V_H Gene Segments Rearranged *in Vitro*

Hennie R. Hoogenboom[1] and Greg Winter[1,2]†

[1]*MRC Centre for Protein Engineering*
and [2]*MRC Laboratory of Molecular Biology*
Hills Road, Cambridge CB2 2QH, U.K.

(Received 13 April 1992; accepted 24 June 1992)

By display of antibody repertoires on the surface of a filamentous bacteriophage and selection of the phage by binding to antigen, we can mimic immune selection. Recently, by tapping the repertoire of rearranged V-genes from the peripheral blood lymphocytes of unimmunised donors, we succeeded in making human antibody fragments with different specificities, including both haptens and proteins, from the same library of phage. Now we have built a repertoire of human V_H genes from 49 human germline V_H gene segments rearranged *in vitro* to create a synthetic third complementarity determining region (CDR) of five or eight residues. The rearranged V_H genes were cloned with a human $V_\lambda 3$ light chain as single chain Fv fragments for phage display, and the library of phage panned by binding to each of two haptens, 2-phenyl-5-oxazolone (phOx) or 3-iodo-4-hydroxy-5-nitrophenyl-acetate (NIP) coupled to bovine serum albumin (BSA). Many different antibody fragments were isolated which bound specifically to hapten, some with affinities in the micromolar range. The *in vitro* "immune response" to the hapten NIP was dominated by the 9-1 segment ($V_H 3$ family), and that to phOx by the VH26 segment ($V_H 3$ family) with an invariant aromatic residue (Tyr, Phe, Trp) at residue 97 of CDR3. However, the isolation of phage against protein antigens proved more elusive, with a single phage binding to human tumour necrosis factor, and none to bovine serum albumin, turkey egg-white lysozyme or human thyroglobulin. Nevertheless, the work shows that human antibody fragments with specific binding activities can be made entirely *in vitro*.

Keywords: human antibodies; filamentous phage; synthetic libraries; V_H segments

The display of antibody fragments on the surface of filamentous phage that encode the antibody genes (McCafferty *et al.*, 1990; Hoogenboom *et al.*, 1991; Breitling *et al.*, 1991; Kang *et al.*, 1991), and the selection of phage by binding to antigen (Parmley & Smith, 1988), offer a powerful means of making antibodies from repertoires of V-genes (Ward *et al.*, 1989; Huse *et al.*, 1989). Thus, rearranged V-genes have been harvested from the mRNA of B-cells of immunised mice (Clackson *et al.*, 1991) and humans (Barbas *et al.*, 1991) using the polymerase chain reaction (Saiki *et al.*, 1985), and the heavy and light chains combined at random (Huse *et al.*, 1989) and cloned into filamentous phage for display. The phage have been selected by panning with antigen and the encoded antibody fragments secreted as soluble fragments from infected bacteria (Skerra & Plückthun, 1988; Better *et al.*, 1988). Due to the increase of mRNA in antigen stimulated B-cells compared with resting B-cells, immunisation contributes to the isolation of heavy and light V-genes that are predisposed to create antigen binding combinations (Persson *et al.*, 1991; Clackson *et al.*, 1991; Hawkins & Winter, 1992).

However, the display of antibodies on phage and selection with antigen mimicks immune selection (Milstein, 1990; Winter & Milstein, 1991), and offers the prospect of making antibodies without immunisation from a single "universal" library of phage. Indeed, we recently derived human antibody fragments with moderate binding affinities ($K_d > 0{\cdot}1\ \mu M$) and high specificities to haptens and protein

† Author to whom all correspondence should be addressed.

0022-2836/92/180381-08 $08.00/0

antigens from a repertoire of rearranged V-genes harvested from the peripheral blood lymphocytes of unimmunised donors (Marks *et al.*, 1991). The binding affinity (K_d) of one fragment could be improved to 1 nM by shuffling the antibody chains with those from a repertoire of somatically mutated V-genes (Marks *et al.*, 1992). We have now tried to generate repertoires of rearranged V-gene segments *in vitro*, in order to make artificial antibodies entirely outside the immune system.

Antibody variable domains consist of a β-sheet framework with three loops of hypervariable sequence or CDRs† (Kabat *et al.*, 1991), and the antigen binding site is shaped by loops from both heavy (V_H) and light (V_λ or V_κ) domains. The loops create antigen binding sites of a variety of shapes, ranging from flat surfaces (Amit *et al.*, 1986) to pockets (Alzari *et al.*, 1990). For human V_H domains, the sequence diversity of the first two CDRs are encoded by a repertoire of about 50 germline V_H segments (Tomlinson *et al.*, 1992). The third CDR is generated from the recombination of these segments with about 30 D and six J segments (Ichihara *et al.*, 1988), and although its sequence is highly variable, it often includes a salt bridge from Asp101 of the loop to Arg94 of the framework (Chothia & Lesk, 1987). The structures and lengths of the first two CDRs are restricted (Chothia & Lesk, 1987; Chothia *et al.*, 1989, 1992), but those of CDR3 differ greatly, with lengths ranging from 4 to 25 residues (Kabat *et al.*, 1991).

For human light chain variable domains, the sequence diversity of the first two CDRs and part of CDR3 are encoded by a repertoire of about 50 human V_κ segments (Meindl *et al.*, 1990) and >10 V_λ segments (Chuchana *et al.*, 1990; Combriato & Klobeck, 1991), but the lengths of CDR3 range only from 6 to 10 residues for V_κ genes and 9 to 13 residues for V_λ genes (Kabat *et al.*, 1991). Thus, due to the extensive sequence and length variation of the V_H CDR3, much of the structural diversity is encoded by the V_H domains, and some V_H domains even have antigen binding activities without a light chain partner (Ward *et al.*, 1989).

We created two libraries of rearranged V_H genes with a synthetic CDR3 of five or eight residues, in combination with a single unmutated V_λ light chain. The rearranged V_λ gene was unmutated and derived from the germline IGLV3S1 V_λ segment (Frippiat *et al.*, 1990) isolated from an antibody fragment binding to BSA (αBSA-3) (Marks *et al.*, 1991). Forty-nine germline V_H segments encoding most of the human V_H repertoire (Tomlinson *et al.*, 1992) were each amplified using the polymerase chain reaction (Saiki *et al.*, 1985) and oligonucleotide primers to introduce a J_H4 segment and a synthetic D-segment of 15 bases of random sequence at the 3′ end of the V_H segment (Figs 1 and 2). The five-residue CDR3 is close to the minimal size for this loop, and does not include the salt bridge from Asp101 of the loop to Arg94 of the framework. The eight-residue loop does include Asp101. The rearranged segments were pooled and cloned (Fig. 3) for phage display with a human $V_\lambda3$ light chain (as described in legends of Figs 1, 2 and 3) as single chain Fv fragments (Huston *et al.*, 1988; Bird *et al.*, 1988), creating two synthetic libraries of 10^7 phage clones each (Table 1).

The synthetic libraries tap only a small fraction of the potential diversity. Thus, compared with the

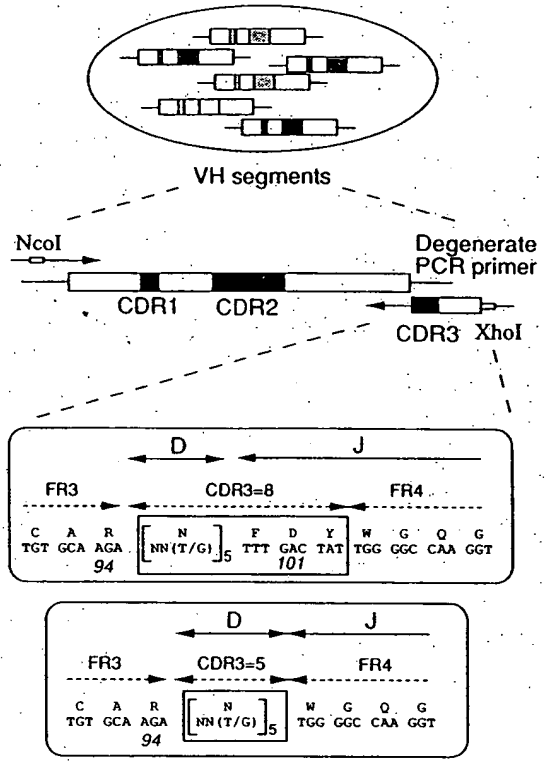

Figure 1. Making rearranged V-genes *in vitro*. A synthetic oligonucleotide (5′ GCC TCC ACC TCT CGA GAC GGT GAC CAG GGT ACC TTG GCC CCA ATA GTC AAA ([A/C]NN)$_5$ TCT TGC ACA GTA ATA CAC GGC CGT GTC 3′) was used to introduce a D-segment with a 5-residue random amino acid sequence, a J-segment and an *Xho*I restriction site, to the 3′ end of each of 49 human V_H germline segments (Tomlinson *et al.*, 1992), creating an 8-residue CDR3. Likewise a synthetic oligonucleotide (5′ GCC TCC ACC TCT CGA GAC GGT GAC CAG GGT ACC TTG GCC CCA ([A/C]NN)$_5$ TCT TGC ACA GTA ATA CAC GGC CGT GTC 3′) was used to create a library with a 5 residue CDR3, in which all 5 residues were random. These primers were used in PCR (Saiki *et al.*, 1985) with V_H family based back primers (VHBACK) incorporating an *Nco*I site (Marks *et al.*, 1991). Each V_H segment clone (provided as single stranded template in M13 vector) was amplified separately at 94 °C for 1 min, 50 °C for 1 min, and 72 °C for 1·5 min, for 25 cycles, on a PHC-3 thermocycler (Techne). Each amplification was checked by electrophoresis on agarose gel, and similar amounts of DNA from V_H segments of the same family were pooled, digested with *Nco*I and *Xho*I, and cloned.

† Abbreviations used: CDR, complementarity determining region; NIP, 3-iodo-4-hydroxy-5-nitrophenylacetate; BSA, bovine serum albumen; phOx, 2-phenyl-5-oxazolone; PCR, polymerase chain reaction.

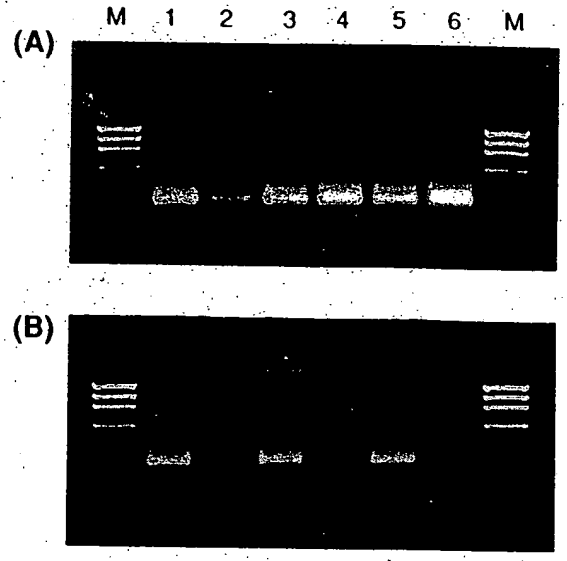

Figure 2. PCR amplification of V_H gene segments. (A) From cloned V_H segments DP-3 (V_H1), DP-26 (V_H2), DP-35 (V_H3), DP-64 (V_H4), DP-73 (V_H5) and DP-74 (V_H6) (Tomlinson *et al.*, 1992), or (B) from lymphocyte DNA from the same donor from whom the V_H genes were cloned, for V_H1, 2, 3, 4, 5 and 6 families, respectively. M, ϕX174 DNA *Hae*III-digested molecular weight marker.

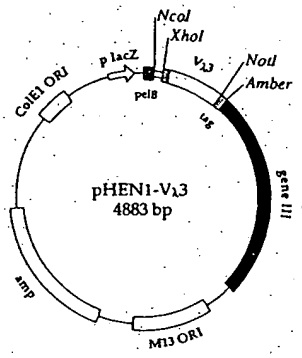

Figure 3. Cloning vector for rearranged V_H genes. A rearranged $V_\lambda3$ light chain variable domain (IGLV3S1; Frippiat *et al.*, 1990) taken from a scFv fragment binding to BSA (Marks *et al.*, 1991) was cloned into the vector pHEN1 (Hoogenboom *et al.*, 1991). The vector pHEN1-$V_\lambda3$ was then cut with *Nco*I and *Xho*I to accept the V_H genes, and encode single chain Fv fragments fused to the phage gene III protein.

library size of 10^7 phage clones, the potential diversity is much greater, corresponding to $49 \times 32^5 = 1{\cdot}6 \times 10^9$ different nucleotide sequences, or $49 \times 20^5 = 1{\cdot}6 \times 10^8$ different amino acid sequences. However, the immune system also uses only a fraction of its potential diversity as there are a limited

Table 1
Composition of the synthetic libraries

Family	No. of genes	DP V_H segments†	Library size $\times 10^{-6}$ (%) CDR3 = 5 (XXXXX)	CDR3 = 8 (XXXXX FDY)
V_H1	14	1–5, 7, 8, 10, 12, 14, 15, 20, 21, 25	2·3 (22)	2·3 (20)
V_H2	1	27	1·0 (10)	1·0 (9)
V_H3	23	29–33, 35, 38–40, 42, 44–54, 58, 59	4·1 (39)	2·1 (19)
V_H4	9	63–71	2·1 (20)	2·6 (23)
V_H5	1	73	0·3 (3)	1·4 (12)
V_H6	1	74	0·6 (6)	1·9 (17)
Total:	49		10·4 (100)	11·3 (100)

Forty nine human V_H segments (Tomlinson *et al.*, 1992) were used, one for each of the V_H2, V_H5 and V_H6 gene families and multiple segments for the other 3 families, and cloned according to family and length of V_H CDR3. Clones from the V_H segments of each family (see legend to Fig. 1) were checked for presence of insert (on average 85%) by PCR-screening (Güssow & Clackson, 1989) with oligonucleotides LMB3 and pHEN-SEQ (Marks *et al.*, 1991) and pooled into 2 large libraries (CDR3 = 5 and CDR3 = 8), creating a bias for certain gene families (the numbers in the Table are corrected for presence of insert). The segments from the V_H2, V_H5, V_H6 families are thereby "over-represented" with respect to the segments from other families. Sequencing double-stranded DNA of 26 and 35 clones from the unselected CDR3 = 5 and CDR3 = 8 libraries, respectively, by the dideoxy chain termination method (Sanger *et al.*, 1977) with oligonucleotide LINKSEQ (5′-CGA TCC GCC ACC GCC AGA G-3′) confirmed that V_H segments from each family were present, and that the nucleotides were present in the expected ratios in the D-segment, but with a slight bias for C. (At the first and second position of each codon, for the CDR3 = 5 library : A, 22·2%; G, 22·1%; C, 31·9% and T, 23·8% ; at the third position, G, 42·3% and T, 57·6% ; for the CDR3 = 8 library : A, 21·3%; G, 17·9%; C, 33·7% and T, 27·1% ; at the third position, G, 42·6% and T, 57·4%.) The expression levels of the soluble scFv antibody fragments were checked by spotting 10 μl supernatant of induced overnight cultures in *E. coli* HB2151 (Hoogenboom *et al.*, 1991) onto a nitrocellulose filter using a slot-blot device (Minifold II, Schleicher and Schuell), and detecting the bound peptide-tagged scFv fragments with 9E10 antibody (Munro & Pelham, 1986) and peroxidase labelled anti-mouse antibodies (Sigma). V_H segments were identified in clones with detectable expression levels, for example V_H1 (DP-7), V_H2 (DP-27), V_H3 (DP-29, 35, 38, 44, 47, 51, 53), V_H4 (DP-63, 69), V_H5 (DP-73) and V_H6 (DP-74).

† V_H segments are listed according to DP nomenclature of Tomlinson *et al.* (1992) for simplicity.

Table 2

Binding activities isolated from the synthetic library

Clone	CDR3 sequence	Family	Germline gene†	Canonical loop structure†	Affinity (μM)
A. *NIP binders*					
αNIP-14	MRGRH FDY	V_H1	DP-8	1-3	
αNIP-2	SRGLA	V_H3	DP-38	1-U	
αNIP-5	SLGVS	V_H3	DP-38	1-U	1·3±0·1
αNIP-11	NGGRH	V_H3	DP-38	1-U	1·5±0·1
αNIP-10	NFGRM	V_H3	DP-38	1-U	1·0±0·1
αNIP-13	SLGLT	V_H3	DP-38	1-U	0·7±0·1
αNIP-16	PFSPY	V_H3	DP-38	1-U	2·0±0·1
αNIP-6	RLPAR FDY	V_H3	DP-45	1-1	
αNIP-12	RLIGR FDY	V_H3	DP-45	1-1	
αNIP-3	KRPTG FDY	V_H3	DP-47	1-3	
B. *phOx binders*					
αOx-31	SNQGI FDY	V_H3	DP-42	1-1	6·7±0·7
αOx-15	LYGVR FDY	V_H3	DP-45	1-1	
αOx-18	LSGVR FDY	V_H3	DP-45	1-1	
αOx-33	AYNYN FDY	V_H3	DP-47	1-3	
αOx-13	LKKYA FDY	V_H3	DP-47	1-3	3·1±0·2
αOx-9	VHRWS FDY	V_H3	DP-47	1-3	
αOx-7	VKQYL FDY	V_H3	DP-47	1-3	
αOx-30	IRWFD FDY	V_H3	DP-47	1-3	
αOx-12	LKRYS FDY	V_H3	DP-47	1-3	
αOx-5	TSKYT FDY	V_H3	DP-47	1-3	
αOx-3	LRSFS FDY	V_H3	DP-47	1-3	
αOx-20	TRAYL FDY	V_H3	DP-47	1-3	
αOx-21	LKGYA FDY	V_H3	DP-47	1-3	
αOx-4	VRSYS FDY	V_H3	DP-47	1-3	
αOx-10	TRRYS FDY	V_H3	DP-47	1-3	
αOx-14	TVGFR FDY	V_H3	DP-47	1-3	
αOx-19	LRRYL FDY	V_H3	DP-47	1-3	
αOx-25	TRAYR FDY	V_H3	DP-47	1-3	
αOx-1	SMGSK FDY	V_H4	DP-67	2-1	
αOx-2‡	SWTWG FDY	V_H4	DP-67	2-1	
C. *TNF binder*					
αTNF-10	QNSGH	V_H3	DP-45	1-1	

Phage were prepared from the mixed libraries by rescue with VCS-M13, and subjected to rounds of panning as in Marks *et al.* (1991). In previous work the scFv fragments were purified using antibody (9E10) directed against a C-terminal tag. Here we used protein A-Sepharose (Pharmacia) as there is a Staphylococcal protein A binding site present on most human V_H3 domains (Sasso *et al.*, 1991). Otherwise the same conditions for loading, washing and elution of fragments were as described previously for the 9E10 antibody (Marks *et al.*, 1991). The binding affinities were measured on samples purified by gel filtration on a Superdex 75 column (Pharmacia). (A) *NIP binders.* Phage binding to NIP-BSA coated tubes were identified by ELISA after 4 rounds of selection with NIP-BSA coated tubes (1 mg/ml in PBS). Of 14 binders sequenced, 10 were unique clones, using 4 germline genes. Germline DP-8 and DP-45 differ only in one or a few amino acid residues from V_{1-2} (Shin *et al.*, 1991) or 65-2 (Matsuda *et al.*, 1990), respectively; DP-38 is identical to 9-1 (Berman *et al.*, 1988); DP-47 is identical to VH26 (Matthyssens & Rabbits, 1980, corrected in Chen *et al.*, 1988). The affinity of 5 clones was measured by fluorescence quench titration (Eisen, 1964) with NIP caproic acid. (B) *phOx binders.* The sequences of 29 phage binding to phOx isolated after 4 rounds of panning on phOx-BSA coated tubes (1 mg/ml in PBS), revealed 20 unique clones using four germline V_H segments, DP-42, 45, 47 (V_H3 family) and DP-67 (V_H4 family). DP-42 and DP-67 only differ in one or a few framework residues from 8-1B (Berman *et al.*, 1988) or $V_H4.22$ (Sanz *et al.*, 1989) respectively. An inhibition ELISA (Rath *et al.*, 1988) was performed with 4-γ-amino-butyric acid methylene 2-phenyl-5-oxazolone (phOx-GABA), with concentrations ranging from 6 to 400 μM in 2% Marvel-PBS. The clones using the DP-47 segment were ranked in the Table by the concentration of phOx-GABA necessary for a 50% reduction of the signal in the inhibition ELISA. The affinities of the clones αOx-13 and αOx-31 for phOx-GABA were determined by fluorescence quench titration after purification. Ideally, the affinity for the phOx-BSA conjugate would have been measured directly, or that for phOx-caproic acid, but phOx-GABA was used here to allow comparison with the hybridoma data of Foote & Milstein (1991). (C) *TNF binder.* After four rounds of panning on TNF coated tubes (10 μg/ml in PBS), about half the clones bound to TNF by ELISA, all of which (8 clones sequenced) had the DP-45 germline V_H with an identical 5-residue CDR3. V_H segments are listed according to DP nomenclature of Tomlinson *et al.* (1992) for simplicity.

† Tomlinson *et al.* (1992); Chothia *et al.* (1992).

‡ Shows V67A mutation in FR3.

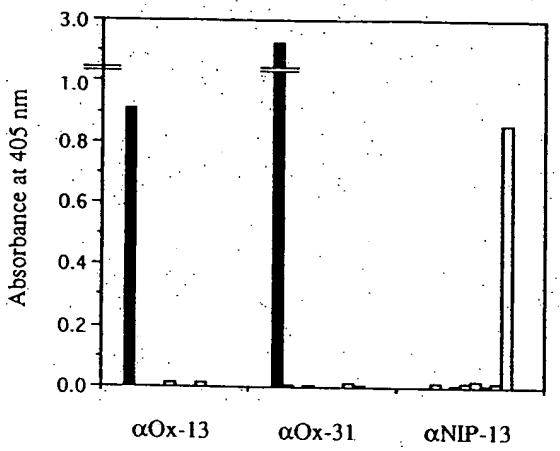

Figure 4. Specificity of binding for the synthetic antibody fragments. Binding was determined by ELISA[4] to phOx-BSA, BSA, cytochrome *c*, turkey egg-white lysozyme, bovine thyroglobulin, chymotrypsinogen A, keyhole limpet haemocyanin, plastic and NIP-BSA (shown left to right), all coated at 100 μg/ml. From each ELISA signal we have subtracted the small background (0·05 to 0·12), corresponding to the signal obtained by incubation of the antigen coated plate with 2 × TY growth medium. All of 20 phOx binders (only 2 phOx binders shown), 10 NIP binders (only 1 NIP binder shown) and 1 TNF binder (not shown) tested bound specifically to phOx-BSA, NIP-BSA and TNF, respectively.

number of B-lymphocytes, for example, about 10^7 to 10^8 for a mouse (Winter & Milstein, 1991), available at any moment in time.

The libraries with five and eight residues in CDR3 were mixed and subjected to four rounds of growth and panning using 2-phenyl-5-oxazolone (phOx)-bovine serum albumin (BSA) and 4-hydroxy-5-iodo-3-nitrophenyl acetyl (NIP)-BSA. The clones were screened as soluble (Hoogenboom *et al.*, 1991) single chain Fv fragments for binding activity to antigen by ELISA (Marks *et al.*, 1991). After the third and fourth rounds, 14/96 and 61/96 clones, respectively, were identified with binding activities to phOx-BSA, and of these (29 tested) none bound to other proteins (Table 2 and Fig. 4). Furthermore, their binding to phOx-BSA coated plates could be competed with the soluble hapten (Table 2, legend). Similarly, NIP-BSA binding clones were isolated after three (2/22) and four (40/96) rounds of panning, and of the 14 binders tested, all bound to NIP-BSA and not to phOx-BSA or BSA, or any of the other proteins used in the specificity ELISA (Fig. 4).

Sequencing revealed that many (20/29) of the phOx binders were unique, with an eight-residue CDR3, and utilised either a segment from the V_H4 family, or one of three segments from the V_H3 family (Table 2). Together, these segments use three of the seven "canonical" folds available to the first two hypervariable loops of human V_H segments (Chothia *et al.*, 1992). The majority of the unique clones (15/20) were derived from the VH26 segment (Matthyssens & Rabbits, 1980; or DP-47, Tomlinson *et al.*, 1992) and have related sequences in the third hypervariable loop: in this group the first residue tends to have a branched aliphatic side-chain (14/15), the second residue tends to be lysine or arginine (11/15), while the fourth residue is always an aromatic residue (most frequently tyrosine). In mouse hybridomas raised against phOx, $V_\kappa$ox1 genes are always, and $V_\kappa$ox-like genes often, found in association with heavy chain genes that encode a short CDR3 with the sequence motif Asp-X-Gly-X-X (where X is any amino acid) (Berek *et al.*, 1985), in which the central glycine creates a cavity for phOx (Alzari *et al.*, 1990). Likewise, the CDR3 motif in the human scFv fragments appears to be a key determinant in binding of phOx, as clones with the VH26 segment and lacking the motif did not bind to phOx, for example, two VH26 clones from the unselected library with CDR sequences Ile-Ile-Pro-Phe-Pro-Phe-Asp-Tyr or Val-Thr-Arg-Ser-Glu-Phe-Asp-Tyr. However, the motif and at least one other CDR of the VH26 segment are required for phOx binding, as other phOx binding clones derived from other V_H gene segments did not use the same sequence motif (Table 2).

Like the phOx binders, the sequences of NIP binders (10 were unique of the 14 sequenced), show a preference for V_H3 family derived germline genes (Table 2). Unlike the phOx binders, both five and eight-residue CDR3 loops, and a V_H1 segment was used. All clones with a five-residue CDR3 used the same germline gene (9-1 or DP-38, Berman *et al.*, 1988; Tomlinson *et al.*, 1992), and in most cases (5/6) had a central glycine, suggesting that, like the murine phOx antibodies, the glycine may help create a binding pocket (see above). The four clones with other germline segments all had eight-residue CDR3 loops, also with a hint of a binding motif, for example αNIP-6 and αNIP-12 with 3/5 residues identical. The same germline V_H gene segments (DP-45 and DP-47) are used by both phOx and NIP binders (but with distinct CDR3s).

The affinities (K_d) of five of the NIP binders (αNIP-13, αNIP-5, αNIP-10, αNIP-11 and αNIP-16) for NIP caproic acid were determined by fluorescence quench titration (Eisen, 1964) and with affinities in the 1 μM range (Table 2) are similar to those of hybridoma antibodies to the hapten NP from a primary mouse immune response to NP (or within an order of magnitude of the affinities of these hybridomas to the heteroclitic NIP; Cumano & Rajewsky, 1986). Similarly, the affinities of two of the phOx binders (αOx-13 and αOx-31, Table 2) for phOx-GABA were slightly poorer than the 1 μM affinities of several hybridomas from a mouse primary immune response to phOx (Foote & Milstein, 1991), but see *caveat*, Table 2, legend. In principle, to improve these affinities, we could systematically alter the antibodies, creating a secondary library for further rounds of selection, for example, by random mutagenesis using PCR (Hawkins *et al.*, 1992), by altering the lengths and sequences of the heavy and light chain CDR3 loops,

or by localising mutations in the other CDR loops, or by chain shuffling (Clackson *et al.*, 1991; Marks *et al.*, 1992).

To determine the influence of the light chain on binding of hapten, the original light chain gene ($V_\lambda 3$ family) of clones αOx-13 and αOx-31 was exchanged for rearranged light chain genes of the $V_\lambda 1$ family (from clone αphOxB2, Marks *et al.*, 1992), $V_\kappa 1$ family (from αTEL9, Marks *et al.*, 1991) or $V_{\kappa 4}$ family (from a clone provided by A.D. Griffiths, unpublished). The binding of the encoded scFv fragments to phOx-BSA is abolished for all the combinations, except for the αOx-13-$V_H/V_\lambda 1$ combination, where there was a weak signal by ELISA. This indicates that the original $V_\lambda 3$ light chain partner binds the antigen directly or helps fix the folding of the CDR loops of the V_H domain. The use of different light chains to construct the primary library would presumably lead to different patterns of V_H gene segments and CDR3 motifs.

Although we generated many different antibody fragments binding to the two haptens, and constructed the synthetic libraries using a light chain derived from a scFv that bound to BSA (with $V_H 3$ segment DP-49 and a CDR3 of 12 residues, Marks *et al.*, 1991), we did not isolate any fragments binding to the carrier protein BSA. Indeed, panning of the library by binding to each of several protein antigens (human tumour necrosis factor (TNF), turkey egg-white lysozyme and human thyroglobulin) provided a single example of phage with antigen-binding activity. After four rounds of panning a phage antibody (and corresponding soluble fragment) was isolated with binding activity to TNF. The fragment (αTNF-10) was derived from the V_H segment DP-45 (Tomlinson *et al.*, 1992). The hapten-binding clones αNIP-6, αNIP-12, αOx-15 and αOx-18 are also derived from this segment (Table 2), although each of these fragments were nevertheless specific for binding to hapten or TNF. This indicates that antigen binding sites with entirely different specificities can be created on the same antibody framework by substitution of CDR3 alone.

The ease in isolating phage that bound to hapten, rather than protein antigen, may be due to the avidity of the antibody fragments on the phage. Although we used a phagemid rather than phage vector for display of the antibody fragments, a significant fraction of the phage population is likely to contain multiple copies of the fusion protein on some phage: furthermore, non-covalently associated dimers of single chain Fv fragments may be present on each head (A. D. Griffiths, J. D. Marks, M. Malmqvist, H. R. Hoogenboom & G. Winter, unpublished data). Due to the higher coating density of hapten on the solid phase, the greater avidity of the phage may selectively favour the binding and retention of hapten binding phage during washing.

Nevertheless, the difficulty in isolating antibody fragments with binding activities to turkey lysozyme (with many potential binding epitopes) contrasts with our isolation of three different antibody fragments to the same antigen from a repertoire of V-genes rearranged *in vivo* (Marks *et al.*, 1991). Model building may yet illuminate whether our choice of light chain and length of CDR sequences might have predisposed the repertoires of structures towards binding of hapten rather than proteins. Antibodies specific for haptens often have deep grooves or pocket-like binding sites, while protein binding antibodies tend to have flat surfaces. Although short V_H CDR3 loops can help create either pockets (Holm *et al.*, 1990) or flat surfaces (de la Paz *et al.*, 1986), the isolation of highly related CDR3 sequences in the synthetic antibodies suggests that our repertoire may have been over-represented for hapten binding sites, at the expense of sites for other antigens. We therefore suspect that this synthetic library, with only short CDR3 lengths of five and eight residues, and confined to a single light chain, poorly recreates the range and distribution of antigen binding shapes required for binding to a large range of antigens. For making synthetic libraries from which antibodies can be isolated against any antigen, we presumably need to create a greater diversity of shapes of antigen binding sites, for example, by increasing both the number of light chains and the diversity of V_H CDR3 lengths. (However, as the CDR3 length increases, the antigen binding may become increasingly dictated by this loop, and in the extreme, the antibody framework may provide little more than a tether for a constrained peptide library.)

Although we have favoured the use of synthetic antibody libraries based on the existing V-gene segments, the thrust of the technology appears to be that an entirely artificial repertoire could serve to cover the antigenic universe. For example, we could build the antigen binding sites on a single human antibody framework (Jones *et al.*, 1986) by introducing loops of diverse or random sequence. In the extreme all six CDRs could be provided by random sequence, although only an infinitesimal fraction of the potential repertoire ($>10^{30}$)(see Winter & Milstein, 1991) could be cloned. Such repertoires could be dominated by structures that do not fold into domains, are not stable or do not associate, so reducing the effective size of the repertoire. Furthermore, the use of loops of random sequence may not create the optimal distribution of binding site structures. By contrast, the loop structures of the V-gene segments form only a limited number of distinct folds and combinations of folds (Chothia *et al.*, 1992) and have presumably evolved for stability and to create a distribution and range of binding sites well matched to the universe of hostile antigens. The V-gene segments may have other desirable features, as the framework regions and first two hypervariable loops of both heavy and light chains of the synthetic human antibodies are likely to be identical in many different individuals (Tomlinson *et al.*, 1992). Such synthetic human antibodies could be less immunogenic in humans than entirely artificial structures.

Finally we must consider whether the synthetic libraries are likely to have advantages over "natural" libraries where the V-genes have been rearranged *in vivo*. A natural library is a black box in which the composition and frequencies of the V-genes are unknown, and among other things, imposed by allelic variation, deletion polymorphism and deletion of self-reactive clones. By contrast, in the synthetic library it is possible to define the sequences and structural elements that may be present, and their frequencies. (For this reason we used 49 defined V_H gene segments to construct the library, rather than repertoires of V-gene segments harvested directly by PCR from the genomic DNA of a donor, as shown in Figure 2.) Although synthetic libraries could be used to recreate a "defined" natural library, they offer the further possibilities of incorporating design into the repertoire of structures. For example, by using the structures of the V-gene segments as building blocks, the repertoire of binding sites might be designed to match the shape of an antigen of known structure. Combining such design with selection offers the prospect of making binding sites of predefined geometry.

We thank I. M. Tomlinson, G. Walter, M. B. Llewelyn and J. D. Marks for human V_H germline genes, J. Foote for phOx-GABA, A. D. Griffiths for help with the fluorescence quench titrations, C. Marks for the D1.3 Fv fragment (as control in the fluorescence quench titrations), T. P. Bonnert for proof-reading, and T. Simon and C. Milstein for discussions. H.R.H. was supported by the D. Collen Research Foundation, Leuven, and the European Molecular Biology Organisation.

References

Alzari, P. M., Spinelli, S., Mariuzza, R. A., Boulot, G., Poljak, R. J., Jarvis, J. M. & Milstein, C. (1990). Three-dimensional structure determination of an anti-2-phenyloxazolone antibody: the role of somatic mutation and heavy/light chain pairing in the maturation of an immune response. *EMBO J.* **9**, 3807–3814.

Amit, A. G., Mariuzza, R. A., Phillips, S. E. & Poljak, R. J. (1986). Three-dimensional structure of an antigen-antibody complex at 2·8 Å resolution. *Science*, **233**, 747–753.

Barbas, C. F., Kang, A. S., Lerner, R. A. & Benkovic, S. J. (1991). Assembly of combinatorial libraries on phage surfaces: the gene III site. *Proc. Nat. Acad. Sci., U.S.A.* **88**, 7978–7982.

Berek, C., Griffiths, G. M. & Milstein, C. M. (1985). Molecular events during maturation of the immune response to oxazolone. *Nature (London)*, **316**, 412–418

Berman, J. E., Mellis, S. J., Pollock, R., Smith, C. L., Suh, H., Heinke, B., Kowal, C., Surti, U., Cantor, C. R. & Alt, F. W. (1988). Content and organization of the human Ig V_H locus: definition of three new V_H families and linkage to the Ig CH locus. *EMBO J.* **7**, 727–738.

Better, M., Chang, C. P., Robinson, R. & Horwitz, A. H. (1988) *Escherichia coli* secretion of an active chimeric antibody fragment. *Science*, **240**, 1041–1043.

Bird, R. E., Hardman, K. D., Jacobson, J. W., Johnson, S., Kaufman, B. M., Lee, S.-M., Lee, T., Pope, S. H., Riordan, G. S. & Whitlow, M. (1988). Single-chain antigen-binding proteins. *Science*, **242**, 423–426.

Breitling, S. D., Seehaus, T., Klewinghaus, I. & Little, M. (1991). A surface expression vector for antibody screening. *Gene*, **104**, 147–153.

Chen, P. P., Liu, M. F., Sinha, S. & Carson, D. A. (1988). A 16/6 idiotype-positive anti-DNA antibody is encoded by a conserved V_H gene with no somatic mutation. *Arthritis Rheum.* **31**, 1429–1431.

Chothia, C. & Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **196**, 901–917.

Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith, G. S., Air, G., Sheriff, S., Padlan, E. A., Davies, D., Tulip, W. R. *et al.* (1989). Conformations of immunoglobulin hypervariable regions. *Nature (London)*, **342**, 877–883.

Chothia, C., Lesk, A., Gherardi, E., Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn, M. B. & Winter, G. (1992). The structural repertoire of the human V_H segments. *J. Mol. Biol.* **227**.

Chuchana, P., Blancher, A., Brockley, F., Alexandre, D., Lefranc, G. & Lefranc, M.P. (1990). Definition of the human immunoglobulin variable lambda (IGLV) gene subgroups. *Eur. J. Immunol.* **20**, 1317–1325.

Clackson, T., Hoogenboom, H. R., Griffiths, A. D. & Winter, G. (1991). Making antibody fragments using phage display libraries. *Nature (London)*, **352**, 624–628.

Combriato, G. & Klobeck, H. G. (1991). Vλ and Jλ-Cλ gene segments of the human λ light chain locus are separated by 14 kb and rearranged by a deletion mechanism. *Eur. J. Immunol.* **21**, 1513–1522.

Cumano, A. & Rajewski, K. (1986). Clonal recruitment and somatic mutation in the generation of immunological memory to the hapten NP. *EMBO J.* **5**, 2459–2468.

de la Paz, P., Sutton, B. J., Darsley, M. J. & Rees, A. R. (1986). Modelling of the combining sites of three anti-lysozyme monoclonal antibodies and of the complex between one of the three antibodies and its epitope. *EMBO J.* **5**, 415–425.

Eisen, H. N. (1964). Determination of antibody affinity for haptens and antigen by means of fluorescence quenching. *Methods Med. Res.* **10**, 115–121.

Foote, J. & Milstein, C. (1991). Kinetic maturation of an immune response. *Nature (London)*, **352**, 530–532.

Frippiat, J. P., Chuchana, P., Bernard, F., Buluwela, L., Lefranc, G. & Lefranc, M. P. (1990). First genomic sequence of a human Ig variable lambda gene belonging to subgroup III. *Nucl. Acids Res.* **18**, 7134.

Güssow, D. & Clackson, T. (1989). Direct clone characterization from plaques and colonies by the polymerase chain reaction. *Nucl. Acids Res.* **17**, 4000.

Hawkins, R. E. & Winter, G. (1992). Cell selection strategies for making antibodies from variable gene libraries: trapping the memory pool. *Eur. J. Immunol.* **22**, 867–870.

Hawkins, R. E., Russell, S. J. & Winter, G. (1992). Phage antibodies: affinity maturation by random mutagenesis and selection. *J. Mol. Biol.* **226**, 889–896.

Holm, L., Laasonen, L., Kaartinen, M., Teeri, T. T. & Knowles, J. K. C. (1990). Molecular modelling study of antigen binding to oxazolone-specific antibodies: the Ox1 idiotype IgG and its mature variant with increased affinity to 2-phenyloxazolone. *Protein Eng.* **3**, 403–409.

Hoogenboom, H. R., Griffiths, A. D., Johnson, K. S., Chiswell, D. J., Hudson, P. & Winter, G. (1991).

Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. *Nucl. Acids Res.* **19**, 4133–4137.

Huse, W. D., Sastry, L., Iverson, S. A., Kang, A. S., Alting, M. M., Burton, D. R., Benkovic, S. J. & Lerner, R. (1989). Generation of combinatorial library of the immunoglobulin repertoire in phage lambda. *Science*, **246**, 1275–1281.

Huston, J. S., Levinson, D., Mudgett, H. M., Tai, M. S., Novotny, J., Margolies, M. N., Ridge, R. J., Bruccoleri, R. E., Haber, E., Crea, R. & Opperman, H. (1988). Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. *Proc. Nat. Acad. Sci., U.S.A.* **85**, 5879–5883.

Ichihara, Y., Matsuoka, H. & Kurosawa, Y. (1988). Organization of human immunoglobulin heavy chain diversity gene loci. *EMBO J.* **7**, 4141–4150.

Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. (1986). Replacing the complementarity-determining regions in human antibody with those from a mouse. *Nature (London)*, **321**, 522–525.

Kabat, E. A., Wu, T. T., Perry, H. M., Gottesman, K. S. & Foeller, C. (1991). *Sequences of Proteins of Immunological Interest*. 5th edit. U.S. Department of Health and Human Services, Public Health Service National Institutes of Health.

Kang, A. S., Barbas, S. F., Janda, K. D., Benkovic, S. J. & Lerner, R. (1991). Linkage of recognition and replication functions by assembling combinatorial antibody Fab libraries along phage surfaces. *Proc. Nat. Acad. Sci., U.S.A.* **88**, 4363–4366.

Marks, J. D., Hoogenboom, H. R., Bonnert, T. P., McCafferty, J., Griffiths, A.D. & Winter, G. (1991). By-passing immunization: human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.* **222**, 581–597.

Marks, J. D., Griffiths, A. D., Malmqvist, M., Clackson, T., Bye, J. M. & Winter, G. (1992). By-passing immunization: building high affinity human antibodies by chain shuffling. *Bio/Technology*, **10**, 779–783.

Matsuda, F., Shin, E. K., Hirabayashi, Y., Nagaoka, H., Yoshida, M. C., Zong, S. Q. & Honjo, T. (1990). Organization of variable region segments of the human immunoglobulin heavy chain: duplication of the D5 cluster within the locus and interchromosomal translocation of variable region segments. *EMBO J.* **9**, 2501–2506.

Matthyssens, G. & Rabbitts, T. H. (1980). Structure and multiplicity of genes for the human immunoglobylin heavy chain variable region. *Proc. Nat. Acad. Sci., U.S.A.* **77**, 6561–6565.

McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. (1990). Phage antibodies: Filamentous phage displaying antibody variable domains. *Nature (London)*, **348**, 552–554.

Meindl, A., Klobeck, H. G., Ohnheiser, R. & Zachau, H. G. (1990) The V kappa gene repertoire in the human germline. *Eur. J. Immunol.* **20**, 1855–1863.

Milstein, C. (1990). The Croonian lecture, 1989. Antibodies: a paradigm for the biology of molecular recognition. *Proc. Roy. Soc. Lond. Biol.* **239** (1294), 1–16.

Munro, S. & Pelham, H. R. B. (1986). An Hsp-like protein in the ER: identity with the 78kd glucose regulated protein and immunoglobulin heavy chain binding protein. *Cell*, **46**, 291–300.

Parmley, S. F. & Smith, G. P. (1988). Antibody-selectable filamentous fd phage vectors: Affinity purification of target genes. *Gene*, **73**, 305–318.

Persson, M. A. A., Caothien, R. H. & Burton, D. R. (1991). Generation of diverse high affinity human monoclonal antibodies by repertoire cloning. *Proc. Nat. Acad. Sci., U.S.A.* **88**, 2432–2436.

Rath, S., Stanley, C. M. & Steward, M. W. (1988). An inhibition enzyme immunoassay for estimating relative antibody affinity and affinity heterogeneity. *J. Immunol. Methods*, **106**, 245–249.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487–491.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Nat. Acad. Sci., U.S.A.* **74**, 5463–5467.

Sanz, I., Kelly, P., Williams, C., Scholl, S., Tucker, P. & Capra, J. D. (1989). The smaller human V_H gene families display remarkably little polymorphism. *EMBO J.* **8**, 3741–3748.

Sasso, E. H., Silverman, G. J. & Mannik, M. (1991). Human IgA and IgG $F(ab')_2$ that bind to *Staphylococcal* protein A belong to the VHIII subgroup. *J. Immunol.* **147**, 1877–1883.

Shin, E. K., Matsuda, F., Nagaoka, H., Fukita, Y., Imai, T., Yokoyama, K., Soeda, E. & Honjo, T. (1991). Physical map of the 3' region of the human immunoglobulin heavy chain locus: clustering of autoantibody-related variable segments in one haplotype. *EMBO J.* **10**, 3641–3645.

Skerra, A. & Plückthun, A. (1988). Assembly of a functional immunoglobulin Fv fragment in *Escherichia coli*. *Science*, **240**, 1038–1040.

Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn, M. B. & Winter, G. (1992). The repertoire of human germline V_H sequences reveals about fifty groups of V_H segments with different hypervariable loops. *J. Mol. Biol.* **227**.

Ward, E. S., Güssow, D., Griffiths, A. D., Jones, P. T. & Winter, G. (1989). Binding activities of a repertoire of single immunoglobulin variable domains secreted from *Escherichia coli*. *Nature (London)*, **341**, 544–546.

Winter, G. & Milstein, C. (1991). Man-made antibodies. *Nature (London)*, **349**, 293–299.

Edited by J. Karn